

*Henning Andersen & Henning Bergenholtz**

Brug af surveys til lingvistiske undersøgelser. Om: Erik Jørgen Hansen/Bjarne Hjorth Andersen: *Et sociologisk værktøj. Introduktion til den kvantitative metode*. København: Reitzel, 2000. (272 sider, kr. 250,-)

1. Mini-surveys

Dagens nyheder, det være sig i avis, radio eller fjernsyn, har så godt som hver dag et indslag med en gengivelse af en eller anden forespørgsel af "tilfældige" forbipasserende. I radio- og tv-nyhederne er antallet af adspurgte tilsyneladende identisk med antallet af interviewede personer, dvs. en 3-4 "tilfældige" personer. Det virker realistisk, det er på naturalistisk vis tilsyneladende taget direkte fra det virkelige liv, men er ikke bare ikke repræsentativt, men sandsynligvis næsten altid helt misvisende i relation til et repræsentativt udvalg af befolkningen. Sådan må man også bedømme resultatet af en rundspørge blandt læserne af Readers Digest, som skulle svare på, om de havde fået deres tabte pung tilbage eller ej. 80% af læserne havde fået deres pung tilbage, men man må formode, at det først og fremmest har været de heldige, der har reageret på spørgsmålet.

På lignende vis kan resultatet af svar i den radioudsendelse sandsynligvis interpreteres, hvor lytterne skulle ringe eller skrive for at sige, om der var for mange fremmedord i nutidens sprog. Dette mente næsten alle lytterne (mere end 96%). Også her må man formode, at det er et specielt udvalg af befolkningen med en overvægt af sprogekritiske lyttere, der reagerede. Lignende indvendinger kan fremføres mod en del af de lingvistiske surveys, som gengives i den lingvistiske litteratur.

* *Henning Andersen & Henning Bergenholtz
Handelshøjskolen i Århus
Fuglesangs Allé 4
DK-8210 Århus V*

Som empirisk basis for lingvistiske undersøgelser vælges i praksis en eller flere af følgende grundmuligheder:

1. sprogkompetence
2. tekstundersøgelse
3. foreliggende lingvistisk litteratur

Vi vil i denne anmeldelsesartikel koncentrere os om den første mulighed, som til dels kan bestå af introspektive angivelser. Disse gengives i den lingvistiske litteratur ofte som almene sandheder, der således ikke kun skal dække lingvistens egen introspektion, men også alle andre sprogbrugeres. Lingvisten fremstår således som den repræsentative stikprøve for hele populationen, idet der ikke gøres opmærksom på den meget begrænsede empiriske basis. Denne form for introspektive generaliseringer, som i virkeligheden er en immuniseringsstrategi, bruges også i en anden variation, når der henvises til imaginære respondenter (informerter) (se hertil Bergenholtz (2000) med flere eksempler i anmeldelsen af Sandberg (1998)):

- Alle informanterne anser denne sætning for korrekt.
- Enhver informant vil anse brugen af dette korrelat for at være fakultativ.
- Enhver informant vil afvise muligheden for at indsætte et korrelat i denne sætning.

Man ved i sådanne tilfælde ikke, om lingvisten virkelig har gennemført en lingvistisk survey; hvilke spørgsmål han har stillet; eller hvor mange hhv. hvilke respondenter han har spurgt.

Lidt mindre immuniserende er man i lingvistiske arbejder, hvor der åbent fortælles, at tre, fem eller tyve respondenter er blevet inddraget. Det drejer sig oftest om studerende, enkelte gange som i de følgende eksempler om nære venner og familien. Man finder heri udsagn som i Augst (1975):

- de fleste informanter ...
- alle informanter ...
- ikke alle informanter ...

Som det fremgår af forordet til Augst (1975) har den pågældende undersøgelse gjort brug af tre respondenter. Det har fx også Colliander (1983), som dog i hvert enkelt tilfælde fortæller om hver enkelt respondents holdning til hver problemstilling. Også i de få undersøgelser, som har

mellem 20 og 315 respondenter, fx Bergenholtz (1983) med 100 respondenter, diskuteres ikke, om de adspurgte studerende på nogen måde kan siges at være repræsentative for hele populationen. Der siges intet om fordelingen mellem kvinder og mænd, mellem respondenter fra forskellige landsdele eller med forskellig uddannelse osv.

Meget taler for, at lingvistikken bør tage hensyn til noget af den viden, som man har oparbejdet inden for socialvidenskaberne. I dansk lingvistisk sammenhæng kunne den nyligt udkomne lærebog i kvantitative sociologiske metoder af Erik Jørgen Hansen og Bjarne Hjorth Andersen være en oplagt mulighed.

2. Et sociologisk værktøj

For at sige det straks og før visse indskrænkninger: Det drejer sig om en velskrevet og tilsvarende letlæselig lærebog, som også og måske især er velegnet til selvstudium. Den har – hvad der ikke gælder for alle håndbøger – et indeks over brugte termer. Særligt fordelagtige er de korte en- eller tosætningsssammenfatninger af de enkelte afsnit, som gives i en egen spalte uden for tekstrammen.

Sociale fænomener, fastslås i første kapitel, interesserer i denne sammenhæng ikke som individuelle tilfælde eller som individuelle årsager for bestemte handlinger. De individuelt adspurgte individer gælder som repræsentanter for bestemte sociale grupper. De tre repræsentanter i de fornævnte lingvistiske undersøgelser vil næppe kunne regnes for at være sådanne repræsentanter, snarere må man her tale om en undersøgelse af tre enkeltindivider. Hvor grænsen mellem en sådan enkeltindividrelateret, en kvalitativ og en kvantitativ undersøgelse ligger, bliver i fremstillingen side 22-23 ikke helt klar for den læser, som ikke allerede har et ret godt kendskab til sociologiske teorier og metoder. Den samme indvending kan gøres mod adskillelsen mellem udredningsarbejde og rigtige surveyundersøgelser. Førstnævnte bærer ofte præg af en mangel på forankring i sociologiske teorier og en stor tilbøjelighed til at fremkomme med netop de resultater, som opdragsgiveren kan have fordel af at kunne henvise til. De minder noget om de mini-surveys og de rent introspektive lingvistiske fremstillinger nævnt i indledningen. Det er dog surveyen i sin videnskabelige fundering, som er tema for håndbogen.

I kapitel 3 betones kraftigt det nødvendige i som ved alle andre forskningsområder at gøre sig sine forudsætninger og sine teoretiske referen-

cerammer klart, inden man går i gang med sin undersøgelse. Bogen giver gode henvisninger til, hvordan dette gøres både teoretisk og praktisk. Sammenhængen mellem hensigten med undersøgelsen og valg af metode betones. Samtidig med de overordnede overvejelser gives en håndværksmæssig tilgang til fremgangsmåden med problemformulering, etablering af teorigrundlag, hypoteseformulering, modelopbygning m.v. Formålet med disse overvejelser er helt klart at få understreget, som der står i sammenfatningerne, at en empirisk kortlægning af fænomenerne er nyttig, men forklaringer på fænomenerne er mere spændende. Kort sagt: Bogen indbyder til at lave undersøgelser, der udmærker sig ved kortlægning af sammenhænge snarere end ved registrering af data.

Bogen giver en referenceramme til klassifikation af de baggrundsvariable, der således skal indgå i en given undersøgelse. Der gives med en historisk og teoretisk tilgang en oversigt over, hvilke klassifikationer af sociale, uddannelsesmæssige, tidsmæssige baggrundsvariable der kan benyttes. For at kunne lave undersøgelsen er det væsentligt, at de teoretiske definitioner på variablene kan omsættes til operationelle, og det vil også sige målbare definitioner. Specielt kapitel 4 er yderst instruktivt med både historiske baggrundsoplysninger om sociale klassifikationer og detaljerede oplysninger om enkeltaspekter, fx med en advarsel mod at sammenblende uddannelse og socialgruppe.

Når en bestemt population skal undersøges (fx alle beboere i Danmark i et bestemt tidsrum eller alle registrerede selvmordsforsøg i år 2001) kan man foretage en totaltælling, hvilket dog i mange tilfælde ikke er praktisk muligt. Man udtager derfor en stikprøve baseret på en kendt udvælgelsessandsynlighed. Mange af de tekniske problemer, der melder sig her, vil ikke være relevante for lingvistiske undersøgelser. Derimod er temaet i kapitel 6 ”spørgsmålformulering og spørgsmålskonstruktion” yderst relevant. Der skelnes mellem to hovedtyper af surveys: interview og enquete. I første type kan man igen skelne mellem besøgs- og telefoninterviewning. Det kan her undre, at forf. ikke diskuterer det problem, der særligt ved besøgsinterview består i, at interviewerens nødvendigvis påvirker den interviewede. Der henvises blot til, at de personer, som foretager dataindsamlingen, skal have en grundig uddannelse, men der gives ingen oplysninger om en sådan uddannelses indhold. Ved enqueten skelnes mellem postenqueten, hvor spørgeskemaet sendes, forsynet med

en frankeret svarkuvert, og gruppeenqueten, hvor skemaerne udleveres til en gruppe, der besvarer spørgsmålene på stedet.

Det meste af fremstillingen er lærerig og overbevisende, men læseren med en sprogvidenskabelig baggrund undrer sig dog over afsnittet om ”videnskabelig sprogpraksis og hverdagsprog”. Det er et vigtigt problem, som dog skæmmes af mangel på kendskab til den relevante sprogvidenskabelige litteratur. De angivne filosofiske henvisninger er muligvis relevante i filosofisk sammenhæng, men ikke i forbindelse med en generel forståelse af hverdagssprogets problematik i relation til fagsprog:

”Hverdagsproget er ekspressivt, før det bliver semantisk (Stenzel 1934). Sprog opstår som lydgestus, ikke som sprog. Det er en misforståelse, at sproget sættes sammen af allerede færdige dele (Løgstrup 1976).” (side 110)

Lignende sprogvidenskabelige uklarheder hhv. svært forståelige udtalelser findes også i afsnittet om ”ords betydningsvidde” (side 123f) og ”sætningens tomme pladser” (side 125 f). De mere praktiske følger af de (uklare) sproglige grundovervejelser er derimod instruktive, fx om dobbelte negationer og om spørgsmålenes længde.

Gennemgangen af elementære dataanalyser i kapitel 7 er ganske vist lidt teknisk, men dog stort set forståelig for en ikkematematiker. Denne forenkling fører til mindre uklarheder. På side 161 forklares begrebet standardafvigelsen som spredningsmål på udmærket ligefrem vis, og den umiddelbare anvendelighed af standardafvigelsen illustreres ved oplysningen om, at 68% af observationerne i et normalfordelt datasæt ligger inden for en afstand af plus/minus 1 standardafvigelse fra gennemsnittet, og at 95% af dem ligger inden for plus/minus 2 standardafvigelser fra gennemsnittet. Det er sikkert kun pædagogiske overvejelser, der har afholdt forfatterne fra at forstyrre læserens billede af (stikprøve)standardafvigelsen, som et slags gennemsnitsmål for de enkelte observationers kvadrerede afvigelser fra gennemsnittet ved at dividere med $n-1$ i stedet for n (stikprøvestørrelsen), hvilket ville have givet den teoretisk korrekte formel for stikprøvestandardafvigelsen. Problemstillingen kunne også have været omgået ved, som det er gjort på side 188 ved omtalen af standardafvigelsen for procentandelen (i stikprøven), at have angivet, at der er tale om en tilnærmet formel.

Denne detalje spiller dog ikke nogen væsentlig rolle i helhedsbedømmelsen. Bortset fra den manglende kendskab til sprogvidenskabens

verden må ”Et sociologisk værktøj” siges at være velskrevet og en sagligt givtig lærebog, ikke bare for sociologistuderende, men også for lingvister, som med fordel kan gøre brug af denne lærebog.

3. Kritik af enkelte lingvistiske surveys

I Galberg Jacobsen (1999) omtales et survey eller en læserafstemning, som forfatteren kalder den. Abonnementerne på Mål og Mæle er blevet bedt om at tage stilling til en række dobbeltformer som fx:

eleverne *fnisede/fniste/fnes* af lærerens forklaring.

alle medarbejdere har fået nye *password/passwords*

og vælge én og kun én form, der efter deres mening skulle have plads i Retskrivningsordbogen. En del af hensigten med undersøgelsen er at finde ud af, hvilken form der vil være fremtidens form. 116 af abonnenterne svarede.

Ved tilrettelæggelsen og gennemførelsen af denne undersøgelse kunne der have været hentet inspiration i ”Et sociologisk værktøj”. Problemet med undersøgelsen er i første omgang afgrænsningen af den population, der ønskes undersøgt, og herefter hvorledes respondenterne udvælges, så det sikrer et repræsentativt udsnit af populationen, der gør det muligt at generalisere fra forholdene i stikprøven til populationen. Det synes ikke at have været overvejet på forhånd, hvilken population der egentlig ønskedes undersøgt: var det den skrivende og læsende del af den danske befolkning (med dansk som modersmål?) eller den ”sprogbevidste” del, der eventuelt kunne være særlig interessant som ”trendsættende?”. Et andet problem er udvælgelsesmetoden. Nøjagtigt som ved de omtalte undersøgelser om tabte penge og brugen af fremmedord kan det være, at man først og fremmest får svar fra en særlig type respondenter.

En god måde til at sikre repræsentativitet i udvælgelsen er tilfældig udvælgelse, som man ligeledes kan læse om i bogens kapitel 5. Et bærende element her er princippet om, at respondenterne ikke selv bestemmer, om de vil udvælges, hvilket de jo gør, hvis de selv har taget initiativ til at svare. Nu kan man selvfølgelig ikke tvinge respondenter til at svare, og der vil derfor i enhver undersøgelse – og derfor også i ovennævnte – være et problem med frafald, der gør det nødvendigt at belyse, om den resulterende stikprøve stadig er repræsentativ for populationen. Man kunne sige, at man kunne leve med den ikke tilfældige udvælgelse af

respondenterne, hvis de 116, der svarede, var repræsentative. Galberg Jacobsen er selv inde på problemstillingen, men hans konklusion om, at de 116 fx aldersmæssigt skulle være repræsentative for den danske befolkning kan næppe følges, da det ved sammenligning med befolkningstallene fra Danmarks Statistik let kan konstateres, at der blandt de 116 er en undervægt af yngre og overvægt af ældre (vel ikke en ønskelig sammensætning, når fremtidens former ønskes belyst). Galberg Jacobsen er godt klar over den manglende repræsentativitet mht. uddannelsesmæssig og social baggrund. Men han kan ikke rigtig finde sine ben i bedømmelsen af dette forholds indvirken på undersøgelsens kvalitet, hvilket dybest set hænger sammen med, at der fra starten ikke er taget stilling til, hvilken population der ønskes undersøgt.

En hel snæver tolkning af den population, der er undersøgt, kunne være Mål og Mæles læsere. Ud fra undersøgelsen kan det konkluderes at langt de fleste fundne forskelle er signifikante, dvs. at de med 95 % sandsynlighed dækker over forskelle, der også findes i den bagvedliggende population. Det gælder dog ikke forskellene *fnisede/fniste*, *balstyrig/balstyrisk* samt *selv om/selvom*. Dertil behøves udregning af konfidensintervaller for forskelle i kvalitative variable med flere svarmuligheder, som ikke er medtaget i ”Et sociologisk værktøj”, men kan findes i fx Mortensen (1987, 187)

I Bergenholtz/Jensen (2000) om inddragelse af informanter ved ord-bogsarbejde omtales en lignende undersøgelse af informanternes stillingtagen til dobbeltformer. Også ved tilrettelæggelsen og gennemførelsen af denne undersøgelse kunne der have været hentet inspiration i ”Et sociologisk værktøj”. I undersøgelsen er inddraget 337 informanter. Populationen synes her at være den danske befolkning - eller underforstået, den del af befolkningen, der kan læse og skrive. Den måde, hvorpå informanterne er inddraget, sikrer ikke umiddelbart et miniaturebillede af befolkningen. De 15 ansvarlige for undersøgelsen har fordelt spørgeskemaer mellem venner og bekendte og i familien. Altså helt klart en udvælgelsesmetode, der forsynder sig mod tilfældighedsprincippet i udvælgelsen og derfor som følge af, at respondenterne i høj grad ”ligner” hinanden, skaber grobund for det værste af alt i sådanne undersøgelser – nemlig systematiske fejl (bias), der kan føre til graverende fejlslutninger om forholdene i populationen. Da der ikke er sket nogen tilfældig udvælgelse, kan det derfor ikke undre, at resultaterne fra de enkelte

ansvarliges venner og familie ikke stemmer overens med resultaterne i den samlede undersøgelse. Grunden til, at en enkelt ansvarligs 22 informanter har en 17:5 præference for *nøs*, mens forholdet i hele undersøgelsen er 177:105 for *nyste*, er ikke, at stikprøven, som der står i artiklen, ikke er stor nok, men at resultaterne er indhentet fra helt forskellige (del)populationer, der adskiller sig på grund af fx geografiske, sociale og aldersmæssige forhold. Samlet set kunne de 337 informanter måske ses som et billede af den læsende og skrivende del af befolkningen. En analyse af grunddata i stil med den analyse, der gennemgås i bogens afsnit ”Bortfald – fra stikprøve til analyseudvalg” side 88, ville kunne give svar på dette spørgsmål.

Forudsat at stikprøven kan betragtes som et repræsentativt udsnit af befolkningen, kan det konkluderes, at langt de fleste af de forskelle mellem dobbeltformerne, der er fundet i ovennævnte undersøgelse, kan tolkes som egentlige præferencer for den hyppigst forekommende form i befolkningen. Følgende forskelle er ikke signifikante: *taknemmelighed/taknemlighed*, *utaknemmelighed/utaknemlighed*, *fleksjob/flexjob*, *flekstid/flextid* samt *flekstidsordning/flextidsordning*

Sammenfattende om de to undersøgelser kan man sige, at et studie af ”Et sociologisk værktøj”, som er udkommet efter offentliggørelse af de to nævnte artikler, ville have ledt iværksætterne til en større præcisering af de populationer, der ønskedes undersøgt, samt til en større fokusering på nødvendigheden af at sikre en tilfældig udvælgelse. Det farlige ved undersøgelserne, der altså på trods af de påviste mangler og de relativt små stikprøvestørrelser dog har fundet en række signifikante forskelle, er, at disse resultater uden forbehold bliver opfattet af tredjemand som gældende, og at omstændighederne omkring udvælgelsen glemmes, så det som omtalt i bogen side 185 ff. igen bliver en sjældent nævnt forudsætning, at resultater fra en stikprøve kun har almen gyldighed, såfremt denne er udvalgt på en sådan måde, så den kan betragtes som et repræsentativt udsnit af den relevante population.

4. Perspektivering

Mængden af omfattende og også repræsentative lingvistiske surveys er ringe. Når det drejer sig om dansk almensprog, kender vi ikke til andre større undersøgelser. De beskrevne lever på flere måder ikke op til de kvalitetsstandarder, som på fortræffelig vis fremstilles i ”Et sociologisk

værktøj”. Når og hvis også lingvister har sat sig ind i en sådan grundviden, kan der måske blive basis for en gennemførelse af repræsentative og videnskabeligt forsvarlige undersøgelser. Man kunne evt. fortsætte med grammatiske og ortografiske tvivlstilfælde, men i princippet vil man til enhver lingvistisk undersøgelse kunne inddrage sproglige surveys som en del af den empiriske basis.

5. Litteratur

- Augst, Gerhard (1975). *Lexikon zur Wortbildung und Morpheminventar*. Bd.1: A-G, Bd.2: H-R, Bd.3: S-Z. Tübingen 1975.
- Bergenholtz, Henning (1983). Grammatik im Wörterbuch: Zur Terminologie und zur empirischen Basis. In *Kopenhagener Beiträge zur germanistischen Linguistik 21*, 70-72.
- Bergenholtz, Henning/Sanne Jensen (2000). Inddragelse af informanter ved ordbogsarbejde. In *LexicoNordica 7*, 149-166.
- Bergenholtz, Henning. Informantenbefragungen zur Auffindung von grammatischen Regeln für Textproduktion, für Textrezeption und für die Erlernung einer Sprache. Reflexionen zu: Bengt Sandberg: *Zum es bei transitiven Verben vor satzförmigem Akkusativobjekt*. Tübingen: Narr, 1998. In *Hermes 25*, 2000, 143-158.
- Colliander, Peter (1983). *Das Korrelat und die obligatorische Extraposition*. (= *Kopenhagener Beiträge zur Germanistischen Linguistik, Sonderband 2*). Kopenhagen.
- Galberg Jacobsen, Henrik (1999). Fra dobbelt til enkelt? Kommentar til en læserafstemning. In *Mål og Måle 22.3*, 16-29.
- Mortensen, Peter S. (1987). *Statistik for økonomer. Repræsentative undersøgelser*. Herning: Systime.
- Sandberg, Bengt (1998). *Zum es bei transitiven Verben vor satzförmigem Akkusativobjekt*. Tübingen: Narr.

